PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07H 21/00, C12Q 1/68, A61K 48/00

A2

(11) International Publication Number: WO 99/65928

(43) International Publication Date: 23 December 1999 (23.12.99)

(21) International Application Number: PCT/US99/13647

(22) International Filing Date: 18 June 1999 (18.06.99)

(30) Priority Data:

60/090,039	19 June 1998 (19.06.98)	US
60/090,040	19 June 1998 (19.06.98)	US
60/090,041	19 June 1998 (19.06.98)	US
60/089,853	19 June 1998 (19.06.98)	US
60/089,997	19 June 1998 (19.06.98)	US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US Not furnished (CIP)
Filed on Not furnished

(71) Applicant *(for all designated States except US)*: GENZYME CORPORATION [US/US]; Metrowest Place, 15 Pleasant Street Connector, Framingham, MA 01701-9322 (US).

(71)(72) Applicants and Inventors: ROBERTS, Bruce, L. [US/US]; 26 Windsor Road, Milford, MA 01757 (US). SHANKARA, Srinivas [US/US]; 24 Stoney Hill Road, Shrewsbury, MA 01545 (US).

(74) Agents: KONSKI, Antoinette, F. et al.; Baker & McKenzie, 660 Hansen Way, Palo Alto, CA 94304 (US).

(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
With declaration under Article 17(2)(a); without abstract; title not checked by the International Searching Authority.

(54) Title: POLYNUCLEOTIDE POPULATION ISOLATED FROM NON-METASTATIC AND METASTATIC BREAST TUMOR TISSUES

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

POLYNUCLEOTIDE POPULATION ISOLATED FROM NON-METASTATIC AND METASTATIC BREAST TUMOR TISSUES

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims priority under 35 U.S.C. § 119(e) to the following U.S. Provisional Application Nos.: 60/090,039; 60/090,040; 60/090,041; 60/089,853; and 60/089,997, each filed June 19, 1998, the contents of which are hereby incorporated by reference into the present disclosure.

TECHNICAL FIELD

This invention is in the field of genetic analysis. Specifically, the invention relates to the isolation of polynucleotides that are differentially expressed in primary or metastatic breast cancer. The compositions and methods of the present invention are particularly useful in diagnoses, prognoses and/or treatment of breast cancer.

BACKGROUND OF THE INVENTION

In spite of numerous advances in medical research, cancer remains the second leading cause of death in the United States. In the industrialized nations, roughly one in five persons will die of cancer. Traditional modes of clinical care, such as surgical resection, radiotherapy and chemotherapy, have a significant failure rate, especially for solid tumors. Failure occurs either because the initial tumor is unresponsive, or because of recurrence due to regrowth at the original site and/or metastases.

Breast cancer is one of the most common cancers and is the third leading cause of death from cancers in the United States with an annual incidence of about 180,200 new cases among women in the United States

during 1997. About 1,400 new cases of breast cancer will be diagnosed in men in 1997. In industrialized nations, approximately one in eight women can expect to develop breast cancer. The overall mortality rate for breast cancer has remained unchanged since 1930. It has increased an average of 0.2% per year, but decreased in women under 65 years of age by an average of 0.3% per year. Preliminary data suggest that breast cancer mortality may be beginning to decrease, probably as a result of increased diagnoses of localized cancer and carcinoma *in situ.* See e.g., Marchant (1994) Contemporary Management of Breast Disease II: Breast Cancer, in: *Obstetrics and Gynecology Clinics of North America* **21**:555–560; and Colditz (1993) *Cancer Suppl.* **71**:1480–1489. An estimated 44,190 deaths (43,900 women, 290 men) in 1997 will occur due to breast cancer. The five-year survival rate for localized breast cancer has increased from 72% in the 1940s to 97% today. If the cancer has spread regionally, however, the rate is 76%, and for women with distant metastases the rate is 20%. Survival after a diagnosis of breast cancer continues to decline beyond five years. Sixty-five percent of women diagnosed with breast cancer survive 10 years and 56% survive 15 years.

Thus, despite an ongoing improvement in our understanding of the disease, breast cancer has remained resistant to medical intervention. Most clinical initiatives are focused on early diagnosis, followed by conventional forms of intervention, particularly surgery and chemotherapy. Such interventions are of limited success, particularly in patients where the tumor has undergone metastasis. There remains a considerable need in the art for developing diagnostic methods to monitor or prognose the progression of the disease. There also exists a pressing need to improve the arsenal of therapies available to provide more precise and more effective treatment in a less invasive way.

Tumor formation is a multi-step process where aberrant cells progressively accrue genetic mutations that confer a growth advantage or survival benefit. For example, cancer cells from metastatic lesions have been found to be more aggressive with respect to their rate of growth and capacity to invade other tissues as compared to cancer cells derived from primary

tumors. It is known that genotypic alterations contribute to the aggressive phenotype of metastatic tumor cells. Due to the vast variability in the nature of the genotypic alterations, the identification of genes preferentially expressed in either non-metastatic breast tumor cells or metastatic breast cells has been difficult. Undoubtly, an exhausted search for such genes have considerable value in both the diagnosis of breast cancer as well as in devising new therapeutic strategies to combat this disease.

DISCLOSURE OF THE INVENTION

The present invention addresses these and certain other deficiencies in the prior art in having isolated and characterized a population of polynucleotides corresponding to genes or transcripts that are differentially expressed or transcribed in either non-metastatic or metastatic breast tumor cells. Transcripts that are overexpressed in the non-metastatic breast tumor such as a primary tumor may encode factors that restrict tumor cell growth such as tumor suppressors, pro-apoptotic factors, inhibitory growth factors or molecules that engage in immune recognition. Transcripts that are preferentially expressed in metastatic tumor tissue may encode factors that augment tumor cell growth or confer a survival benefit such as oncogenes, stimulatory growth factors, anti-apoptotic factors or immunosuppressive factors. These populations of polynucleotides associated with the non-metastatic or metastatic state of a breast cell are particularly useful in the diagnoses and the development of therapeutics for metastatic breast cancer.

Accordingly, the present invention provides a method for aiding in the diagnoses of the metastatic condition of a breast cell by determining differential expression of a polynucleotide that is associated with breast cancer progression. In one aspect, the differential expression is characterized by over expression of a polynucleotide having the sequence selected from the group set forth in Table 1, or the encoded polypeptide. In another aspect, the differential expression is characterized by under-expression of a polynucleotide having the sequence selected from the group set forth n Table 2, or the encoded polypeptide.

Another embodiment of the invention is a screen for a potential therapeutic agent that modulates the expression of a polynucleotide associated with the metastatic condition of a breast tumor cell. The method involves contacting a cell with an effective amount of a potential agent, and assaying for a change in expression level of a polynucleotide selected from the group identified in Tables 1 and 2, wherein a change in the expression level is indicative of a candidate therapeutic agent. The potential therapeutic agent can be, but is not limited to, an antisense oligonucleotide, a ribozyme, a ribozyme derivative, an antibody, a liposome, a small molecule, or an inorganic compound.

Yet another embodiment of the invention is a method of reversing the metastatic condition of a breast cell, wherein the cell is characterized by differential expression of polynucleotides of the invention. In the method, a cell is contacted with an agent identified by the above-mentioned method. Still yet another embodiment of the invention is a method of modulating the genotype and/or phenotype of a breast cell by introducing the cell a polynucleotide of the present invention. In one embodiment a polynucleotide or regulatory sequence identified to inhibit the metastatic potential of the tumor cell is introduced into the cell.

The present invention also provides isolated polynucleotides and populations of the isolated polynucleotides that identify a non-metastatic or a metastatic breast tumor cell. The polynucleotides are intended to include DNA, cDNA, RNA and genomic DNA. Expression systems, including gene delivery vehicles such as liposomes, plasmids and viral vectors, and host cells containing the polynucleotides are further provided by this invention.

Further provided are promoter sequences derived from the tags represented in either of Tables 1 or 2.

Additionally, the invention includes nucleic acid probes and primers that hybridize to invention polynucleotides, as well as isolated nucleic acids comprising novel, expressed gene sequences containing these polynucleotides. The present invention also provides polypeptides and proteins encoded by the polynucleotides.

The present invention further provides antisense oligonucleotides, antibodies, hybridoma cell lines and compositions containing the same.

Further provided are polynucleotides that correspond to regulatory sequence to enhance or inhibit of downstream polynucleotides. The regulatory sequences can be inserted upstream of polynucleotides encoding therapeutic genes.

Also provided are databases of sequences cataloging polynucleotides differentially expressed in non-metastatic or metastatic breast cells and methods of using the sequences to identify and analyze genes expressed in a test cell. In one aspect, the sequences are downregulated in a metastatic breast cell and comprises at least one polynucleotide selected from the group identified in Table 2, and their respective complements in a computer readable form. In another aspect, the database of sequences characterizes a metastatic breast cell and contains at least one polynucleotide selected from the group identified in Table 1, and their respective complements in a computer readable form.

BRIEF DESCRIPTION OF THE SEQUENCE LISTING

Sequence ID Numbers 1 through 3175 depict the tags corresponding to distinct transcripts that are preferentially transcribed in the metastatic breast tumor tissue.

Sequence ID Numbers 3176 through 5911 depict the tags corresponding to distinct transcripts that are preferentially transcribed in the primary or non-metastatic breast tumor tissue.

MODE(S) FOR CARRYING OUT THE INVENTION

Throughout this disclosure, various publications, patents and published patent specifications are referenced by an identifying citation. The disclosures of these publications, patents and published patent specifications are hereby incorporated by reference into the present disclosure to more fully describe the state of the art to which this invention pertains.

Definitions

The practice of the present invention will employ, unless otherwise indicated, conventional techniques of immunology, molecular biology, microbiology, cell biology and recombinant DNA. These methods are described in the following publications. See, *e.g.*, Sambrook et al., MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); "PCR: A PRACTICAL APPROACH" (M. MacPherson et al., IRL Press at Oxford University Press (1991)); PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)); ANTIBODIES, A LABORATORY MANUAL (Harlow and Lane, eds. (1988)); and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

As used in the specification and claims, the singular form "a", "an" and "the" include plural references unless the context clearly dictates otherwise. For example, the term "a cell" includes a plurality of cells, including mixtures thereof.

The term "comprising" is intended to mean that the compositions and methods include the recited elements, but not excluding others. "Consisting essentially of" when used to define compositions and methods, shall mean excluding other elements of any essential significance to the combination. Thus, a composition consisting essentially of the elements as defined herein would not exclude trace contaminants from the isolation and purification method and pharmaceutically acceptable carriers, such as phosphate buffered saline, preservatives, and the like. "Consisting of" shall mean excluding more than trace elements of other ingredients and substantial method steps for administering the compositions of this invention. Embodiments defined by each of these transition terms are within the scope of this invention.

The terms "polynucleotide" and "oligonucleotide" can be used interchangeably, and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof.

Polynucleotides may have any three-dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: a gene or gene fragment, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. A polynucleotide may comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component.

The polynucleotides can be both double- and single-stranded molecules. Unless otherwise specified or required, any embodiment of the invention described herein that is a polynucleotide encompasses both the double-stranded form and each of two complementary single-stranded forms known or predicted to make up the double-stranded form.

A "gene" refers to a polynucleotide containing at least one open reading frame that is capable of encoding a particular protein after being transcribed and translated.

A "gene product" refers to the amino acid (*e.g.*, peptide or polypeptide) generated when a gene is transcribed and translated.

As used herein a second polynucleotide "corresponds to" another (a first) polynucleotide if it is related to the first polynucleotide by any of the following relationships:

1) The second polynucleotide comprises the first polynucleotide and the second polynucleotide encodes a gene product.
2) The second polynucleotide is 5' or 3' to the first polynucleotide in cDNA, RNA, genomic DNA, or fragment of any of these polynucleotides. For example, a second polynucleotide may be a fragment of a gene that includes the first and second polynucleotides. The first and second polynucleotides are related in

that they are components of the gene coding for a gene product, such as a protein or antibody. However, it is not necessary that the second polynucleotide comprises or overlaps with the first polynucleotide to be encompassed within the definition of "corresponding to" as used herein. For example, the first polynucleotide may be a fragment of a 3' untranslated region of the second polynucleotide, for example a promoter sequence. The first and second polynucleotide may be fragment of a gene coding for a gene product. The second polynucleotide may be an exon of the gene while the first polynucleotide may be an intron of the gene.

3) The second polynucleotide is the complement of the first polynucleotide.

The "genotype" of a cell refers to the genetic makeup of the cell and/or its gene expression profile. Modulation of the genotype of a cell can be achieved by introducing additional DNA or RNA either as episomes or as an integral part of the chromosomal DNA of the recipient cell. The genotype can also be modulated by altering the expression level, e.g. mRNA abundance, of a particular gene using agents that regulate gene expression.

A "sequence tag" or "tag" or "SAGE tag" is a short sequence, generally under about 20 nucleotides, that occurs in a certain position in messenger RNA. The tag can be used to identify the corresponding transcript and gene from which it was transcribed. A "ditag" is a dimer of two sequence tags.

A "database" denotes a set of stored data which represent a collection of sequences including nucleotide and peptide sequences, which in turn represent a collection of biological reference materials.

A "probe" is any biochemical labeled with radioactive isotopes or tagged in other ways for ease in identification. A probe is used to identify or isolate a gene, a gene product, or a protein. Examples of probes include, but are not limited to, a radioactive mRNA hybridizing with a single strand of its DNA gene, a DNA or cDNA hybridizing with its complementary region in a chromosome, or a monoclonal antibody combining with a specific protein.

A "promoter" is a region on a DNA molecule to which an RNA polymerase binds and initiates transcription. In an operon, the promoter is usually located at the operator end, adjacent but external to the operator. The nucleotide sequence of the promoter determines both the nature of the enzyme that attaches to it and the rate of RNA synthesis.

A "primer" is a short polynucleotide, generally with a free 3' -OH group, that binds to a target or "template" potentially present in a sample of interest by hybridizing with the target, and thereafter promoting polymerization of a polynucleotide complementary to the target.

The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non-amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" refers to either natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics.

As used herein, the term "isolated" means separated from constituents, cellular and otherwise, in which the polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, are normally associated with in nature. As is apparent to those of skill in the art, a non-naturally occurring the polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, does not require "isolation" to distinguish it from its naturally occurring counterpart. In one embodiment, an "isolated" polynucleotide is separated from the 5' and 3' non-coding but contiguous sequences with which it is normally associated with in nature. In addition, a "concentrated", "separated" or "diluted" polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, is distinguishable from its naturally occurring counterpart in that the concentration or number of molecules per volume is greater than "concentrated" or less than "separated" than that of its naturally occurring

counterpart. A polynucleotide, peptide, polypeptide, protein, antibody, or fragments thereof, which differs from the naturally occurring counterpart in its primary sequence or for example, by its glycosylation pattern, need not be present in its isolated form since it is distinguishable from its naturally occurring counterpart by its primary sequence, or alternatively, by another characteristic such as glycosylation pattern. Thus, a non-naturally occurring polynucleotide is provided as a separate embodiment from the isolated naturally occurring polynucleotide. A protein produced in a bacterial cell is provided as a separate embodiment from the naturally occurring protein isolated from a eucaryotic cell in which it is produced in nature.

As used herein, "expression" refers to the process by which polynucleotides are transcribed into mRNA and/or the process by which the transcribed mRNA (also referred to as "transcript") is subsequently being translated into peptides, polypeptides, or proteins. The transcripts and the encoded polypeptides are collectedly referred to as gene product. If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in an eukaryotic cell.

"Differentially expressed" or "differential expression", as applied to nucleotide sequence or polypeptide sequence in a cell or a tissue, refers to overexpression or underexpression of that polynucleotide when compared to that expressed in a control cell or tissue. Underexpression also encompasses absence of expression of a particular polynucleotide as evidenced by the absence of detectable expression in a tested sample when compared to a control. The selection of the appropriate control cell or tissue is dependent on the sample cell or tissue initially selected and the phenotype of the sample that is under investigation. For instance, if the sample cell is a non-metastatic cell derived from a primary tumor, one or more counterparts or metastatic cells of the sample cell can be used as control cells. Counterparts would include, for example, cell lines established from the same or related cells to those found in the sample cell population. For example, the control cell can be any of a counterpart benign cell type, a counterpart non-metastatic cell type.

A gene or transcript is associated with "breast cancer progression" if it yields transcription or translation products at a substantially altered level or in a substantially altered form in cells derived from metastatic breast tumor tissues as compared with cells of a control tissue, and which may play a role in breast tumor metastasis. The gene or transcript can be a normally quiescent gene that becomes activated (such as a dominant cancer-causing gene); it may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes mutated to produce a variant phenotype; it may be a gene that becomes expressed at an abnormally low level (such as a cancer suppresser gene); or it may be a gene exhibiting differential expression, in which the differential expression correlates with tumor metastasis.

A "polymerase chain reaction" ("PCR") is a reaction in which replicate copies are made of a target polynucleotide using a "pair of primers" or a "set of primers" consisting of an "upstream" and a "downstream" primer, and a catalyst of polymerization, such as a DNA polymerase, and typically a thermally-stable polymerase enzyme. Methods for PCR are well known in the art, and taught, for example in MacPherson et al., (1991) and (1995), *supra*. All processes of producing replicate copies of a polynucleotide, such as PCR or gene cloning, are collectively referred to herein as "replication." A primer can also be used as a probe in hybridization reactions, such as Southern or Northern blot analyses.

"Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step in a more extensive process, such as the initiation of a PCR reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

Hybridization reactions can be performed under conditions of different "stringency". In general, a low stringency hybridization reaction is carried out at about 40 °C in 10 X SSC or a solution of equivalent ionic strength/temperature. A moderate stringency hybridization is typically performed at about 50 °C in 6 X SSC, and a high stringency hybridization reaction is generally performed at about 60 °C in 1 X SSC.

When hybridization occurs in an antiparallel configuration between two single-stranded polynucleotides, the reaction is called "annealing" and those polynucleotides are described as "complementary". A double-stranded polynucleotide can be "complementary" or "homologous" to another polynucleotide, if hybridization can occur between one of the strands of the first polynucleotide and the second. "Complementarity" or "homology" (the degree that one polynucleotide is complementary with another) is quantifiable in terms of the proportion of bases in opposing strands that are expected to form hydrogen bonding with each other, according to generally accepted base-pairing rules. A polynucleotide that is 100% complementary to a second polynucleotide is understood to be "complements" of each other.

"Tumor" or "cancer" comprises a localized population of proliferating cells in an animal that are not governed by the usual limitation of normal growth. The tumor is said to be benign if it does not undergo metastasis and malignant if it undergoes metastasis. A metastatic cell or tissue means that the cell can invade and destroy neighboring body structures.

A "composition" is intended to mean a combination of active agent and another compound or composition, inert (for example, a detectable agent or label) or active, such as an adjuvant.

A "pharmaceutical composition" is intended to include the combination of an active agent with a carrier, inert or active, making the composition suitable for diagnostic or therapeutic use *in vitro*, *in vivo* or *ex vivo*.

As used herein, the term "pharmaceutically acceptable carrier" encompasses any of the standard pharmaceutical carriers, such as a phosphate

buffered saline solution, water, and emulsions, such as an oil/water or water/oil emulsion, and various types of wetting agents. The compositions also can include stabilizers and preservatives. For examples of carriers, stabilizers and adjuvants, see Martin, REMINGTON'S PHARM. SCI., 15th Ed. (Mack Publ. Co., Easton (1975)).

An "effective amount" is an amount sufficient to effect beneficial or desired results. An effective amount can be administered in one or more administrations, applications or dosages.

A "subject," "individual" or "patient" is used interchangeably herein, which refers to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets.

A "control" is an alternative subject or sample used in an experiment for comparison purpose. A control can be "positive" or "negative". For example, where the purpose of the experiment is to determine a correlation of an altered expression level of a gene with a particular type of cancer, it is generally preferable to use a positive control (a subject or a sample from a subject, carrying such alteration and exhibiting syndromes characteristic of that disease), and a negative control (a subject or a sample from a subject lacking the altered expression and clinical syndrome of that disease).

A "gene delivery vehicle" is defined as any molecule that can carry inserted polynucleotides into a host cell. Examples of gene delivery vehicles are liposomes, viruses, such as baculovirus, adenovirus and retrovirus, bacteriophage, cosmid, plasmid, fungal vectors and other recombination vehicles typically used in the art which have been described for expression in a variety of eukaryotic and prokaryotic hosts, and may be used for gene therapy as well as for simple protein expression.

A "viral vector" is defined as a recombinantly produced virus or viral particle that comprises a polynucleotide to be delivered into a host cell, either *in vivo*, *ex vivo* or *in vitro*. Examples of viral vectors include retroviral vectors, adenovirus vectors, adeno-associated virus vectors and the like. In aspects where gene transfer is mediated by a retroviral vector, a vector

construct refers to the polynucleotide comprising the retroviral genome or part thereof, and a therapeutic gene. As used herein, "retroviral mediated gene transfer" or "retroviral transduction" carries the same meaning and refers to the process by which a gene or nucleic acid sequences are stably transferred into the host cell by virtue of the virus entering the cell and integrating its genome into the host cell genome. The virus can enter the host cell via its normal mechanism of infection or be modified such that it binds to a different host cell surface receptor or ligand to enter the cell. As used herein, retroviral vector refers to a viral particle capable of introducing exogenous nucleic acid into a cell through a viral or viral-like entry mechanism.

Retroviruses carry their genetic information in the form of RNA; however, once the virus infects a cell, the RNA is reverse-transcribed into the DNA form which integrates into the genomic DNA of the infected cell. The integrated DNA form is called a provirus.

In aspects where gene transfer is mediated by a DNA viral vector, such as an adenovirus (Ad) or adeno-associated virus (AAV), a vector construct refers to the polynucleotide comprising the viral genome or part thereof, and a therapeutic gene. Adenoviruses (Ads) are a relatively well characterized, homogenous group of viruses, including over 50 serotypes (see, *e.g.*, WO 95/27071). Ads are easy to grow and do not require integration into the host cell genome. Recombinant Ad-derived vectors, particularly those that reduce the potential for recombination and generation of wild-type virus, have also been constructed (see, WO 95/00655; WO 95/11984). Wild-type AAV has high infectivity and specificity integrating into the host cells genome. (Hermonat and Muzyczka (1984) *PNAS USA* **81**:6466-6470; Lebkowski et al. (1988) *Mol. Cell. Biol.* **8**:3988-3996).

Vectors that contain both a promoter and a cloning site into which a polynucleotide can be operatively linked are well known in the art. Such vectors are capable of transcribing RNA *in vitro* or *in vivo*, and are commercially available from sources such as Stratagene (La Jolla, CA) and Promega Biotech (Madison, WI). In order to optimize expression and/or *in vitro* transcription, it may be necessary to remove, add or alter 5' and/or 3'

untranslated portions of the clones to eliminate extra, potential inappropriate alternative translation initiation codons or other sequences that may interfere with or reduce expression, either at the level of transcription or translation. Alternatively, consensus ribosome binding sites can be inserted immediately 5' of the start codon to enhance expression.

Gene delivery vehicles also include several non-viral vectors, including DNA/liposome complexes, and targeted viral protein DNA complexes. Liposomes that also comprise a targeting antibody or fragment thereof can be used in the methods of this invention. To enhance delivery to a cell, the nucleic acid or proteins of this invention can be conjugated to antibodies or binding fragments thereof which bind cell surface antigens, e.g., TCR, CD3 or CD4.

"Host cell" is intended to include any individual cell or cell culture which can be or have been recipients for vectors or the incorporation of exogenous polynucleotides, polypeptides and/or proteins. It also is intended to include progeny of a single cell, and the progeny may not necessarily be completely identical (in morphology or in genomic or total DNA complement) to the original parent cell due to natural, accidental, or deliberate mutation. The cells may be procaryotic or eucaryotic, and include but are not limited to bacterial cells, yeast cells, plant cells, insect cells, animal cells, and mammalian cells, e.g., murine, rat, simian or human.

An "antibody" is an immunoglobulin molecule capable of binding an antigen. As used herein, the term encompasses not only intact immunoglobulin molecules, but also anti-idiotypic antibodies, mutants, fragments, fusion proteins, humanized proteins and modifications of the immunoglobulin molecule that comprise an antigen recognition site of the required specificity. The specificity of an antibody refers to the ability of the antibody to distinguish polypeptides comprising the immunizing epitope from other polypeptides.

As used herein, "solid phase support" is not limited to a specific type of support. Rather a large number of supports are available and are known to one of ordinary skill in the art. Solid phase supports include silica gels, resins,

derivatized plastic films, glass beads, cotton, plastic beads. alumina gels. A suitable solid phase support may be selected on the basis of desired end use and suitability for various synthetic protocols. For example, for peptide synthesis, solid phase support may refer to resins such as polystyrene (*e.g.*, PAM-resin obtained from Bachem Inc., Peninsula Laboratories, etc.), POLYHIPE® resin (obtained from Aminotech, Canada), polyamide resin (obtained from Peninsula Laboratories), polystyrene resin grafted with polyethylene glycol (TentaGel®, Rapp Polymere, Tubingen, Germany) or polydimethylacrylamide resin (obtained from Milligen/Biosearch, California). In a preferred embodiment for peptide synthesis, solid phase support refers to polydimethylacrylamide resin.

The phenotype of a cell is determined by the genes expressed within it. The total of expressed genes can be identified by the "transcripts" (transcribed genes represented by the mRNA population) present in the cell. The totality of transcripts present in any particular cell, affected by certain environmental factors or stimuli, and with varying levels of expression of various transcripts in the cell, can be represented by a "transcriptome". The transcriptome is one means by which to identify the cell.

Serial Analysis of Gene Expression or "SAGE" (Velculescu, et al. (1995) *Science* **270**:484-487 and U.S. Patent No. 5,695,937), provides the tool by which the expressed genes and the expression level of the genes of a cell at any one point in the cell cycle and under various environmental stimuli are isolated, sequenced and cataloged. SAGE provides quantitative gene expression data without the prerequisite of a hybridization probe for each transcript. SAGE is based on two principles. First, a short sequence tag (9-11 base pairs) contains sufficient information to uniquely identify a transcript, provided that it is derived from a defined location within that transcript. Second, many transcript tags can be concatenated into a single molecule and then sequenced, revealing the identity of multiple tags simultaneously. The expression pattern of any population of transcripts can be quantitatively evaluated by determining the abundance of individual tags and

identifying the gene corresponding to each tag. Velculescu et al. (1995) *supra* at 484.

Primary and metastatic breast tumor tissue from the same individual has been subjected to SAGE and the tags isolated from each population were compared and analyzed. Therapeutic relevant tags have been isolated. The polynucleotides comprising or corresponding to these tags, as well as polypeptides and antibodies thereto, are aspects of the present invention.

Polynucleotides, Vectors and Host Cells of the Invention

The present invention provides a polynucleotide and populations of polynucleotides that are differentially expressed in a non-metastatic breast tumor as compared to a metastatic breast tumor, or vice versa. The populations of polynucleotides are characterized in whole or in part by the tags represented in Tables 1 and 2, below, or their respective complements. A polynucleotide is determined to be differentially expressed in a non-metastatic breast tumor cell if it is "overexpressed" or "underexpressed" at least 3 fold higher or less the same or corresponding polynucleotide in the metastatic counterpart. In one embodiment, the population of polynucleotides contains tags corresponding to transcripts that are overexpressed in cells derived from a primary breast tumor. In another embodiment, the population of polynucleotides contains tags or transcripts that are overexpressed in cells derived from a metastatic breast tumor. In further embodiments, the transcript or gene has been previously characterized, but was heretofore unknown to be differentially expressed in a metastatic or a non-metastatic breast tumor tissue. These genes or transcripts can be identified, in whole or in part, by specifically hybridizing under moderate or stringent conditions to the polynucleotides comprising or corresponding to polynucleotides identified in Tables 1 and 2, or their respective complements, using the methods described below.

This invention also provides several embodiments comprising different populations identified by the Sequence ID Nos. as follows: 1, 1-5, 1-17, 18-24, Nos. 1-24, 25-36, 1-36, 18-36, 37-53, 54-74, 37-74, 1-53, 1-74, 75-116, 1-116, 117-279, 1-279, 280-549, 1-549, 550-1160, 1-1160, 1161-3175, 1-3175,

3176-3183, 3184-3197, 3176-3197, 3198-3204, 3176-3204, 3205-3213, 3176-3213, 3214-3226, 3176-3226, 3227-3242, 3176-3242, 3243-3294-3176-3294, 3295-3381, 3176-3381, 3382-3554, 3176-3354, 3555-4012, 3176-4012, 4013-5911-3176-5911, 1-5911, or any combination thereof.

In a separate embodiment, the genes or transcripts are identified using sequence homology or allignment software and sequence databases, as described below.

Hybridization can be performed under conditions of different "stringency". Conditions that vary levels of stringency are well known in the art. See, for example, Sambrook, et al. *supra*. Briefly, relevant conditions include temperature, ionic strength, time of incubation, the presence of additional solutes in the reaction mixture such as formamide, and the washing procedure. Higher stringency conditions are those conditions, such as higher temperature and lower sodium ion concentration, which require higher minimum complementarity between hybridizing elements for a stable hybridization complex to form. In general, a moderate stringency hybridization is typically performed at about 50 ^{0}C in 6 X SSC, and a high stringency hybridization reaction is generally performed at about 60 ^{0}C in 1 X SSC.

A number of the polynucleotide sequences disclosed herein are "novel", that is, the tag or its respective complement, lacks substantial sequence homology with any previously identified Expressed Sequence Tags ("EST") or characterized gene sequences. The inventors have searched databases and if no match is found, the "Description" column is blank indicating that no tag has been identified. If the tag corresponds to an EST or gene, the accession number and/or description of the gene or its product are provided in the Tables.

Additional sequence homology searches can be made with the aid of computer methods. A variety of software programs are available in the art. Non-limiting examples of these programs are Blast (Blast is available from the worldwide web at http://www.ncbi.nlm.nih.gov/BLAST/), DNA Star, MegAlign, and GeneJocky. Any sequence database that contains DNA or

protein sequences corresponding to a gene or a segment thereof can be used for sequence analysis. Commonly employed databases include but are not limited to GenBank, EMBL, DDBJ, PDB, SWISS-PROT, EST, STS, GSS, and HTGS. Sequence similarity can be discerned by aligning the tag sequence against a DNA sequence database. Alternatively, the tag sequence can be translated into six reading frames; the predicted peptide sequences of all possible reading frames are then compared to individual sequences stored in a protein database. Parameters for determining the extent of homology set forth by one or more of the aforementioned alignment programs are well established in the art. They include but are not limited to p value and percent sequence identity. P value is the probability that the alignment is produced by chance. For a single alignment, the p value can be calculated according to Karlin et al. (1990) *Proc. Natl. Acad. Sci* **87**: 2246. For multiple alignments, the p value can be calculated using a heuristic approach such as the one programmed in Blast. Percent sequence identify is defined by the ratio of the number of nucleotide or amino acid matches between the query sequence and the known sequence when the two are optimally aligned. A tag sequence is considered to lack substantial homology with any known sequences when the regions of alignment of comparable length exhibit less than 30% of sequence identity, more preferably less than 20% identity, even more preferably less than 10% identity.

The polynucleotides embodied in the present invention also include larger fragments or the full length coding sequences that comprise a novel sequence identified in Tables 1 and 2. Based on the novel sequences disclosed herein, fragments or the full length coding sequences of the corresponding novel transcripts or genes can be identified using various cloning methods known to artisans in the art. Five methods are disclosed in the section "Methods of Cloning Novel Transcripts or Genes" which further assist practitioners of ordinary skill to isolate these transcripts, genes or cDNA containing or corresponding to the tag sequences of the invention.

In addition to the sequences shown in Tables 1 and 2, this invention also provides the anti-sense polynucleotide stand, e.g. antisense RNA to these

sequences or their complements. One can synthesize an antisense RNA based on the sequences provided in the Tables using any methods available in the art, such as the methodology described in Vander Krol et al. (1988) *BioTechniques* 6:958.

The invention also encompasses polynucleotides which differ from that of the polynucleotides described above, but encode substantially the same amino acid sequences. These altered, but phenotypically equivalent polynucleotides are referred to as "functionally equivalent nucleic acids." As used herein, "functionally equivalent nucleic acids" encompass nucleic acids characterized by slight and non-consequential sequence variations that will function in substantially the same manner to produce the same protein product(s) as the nucleic acids disclosed herein (e.g. by virtue of the degeneracy of the genetic codes), or that have conservative amino acid variations. For example, conservative variations include substitution of a non-polar residue with another non-polar residue, or substitution of a charged residue with a similarly charged residue. These sequence variations include those recognized by artisans in the art as those that do not substantially alter the tertiary structure of the encoded protein.

The polynucleotides of the invention can comprise and can be used to identify additional sequences, such as additional encoding sequences within the same transcription unit, controlling elements such as promoters, ribosome binding sites, and polyadenylation sites, additional transcription units under control of the same or a different promoter, sequences that permit cloning, expression, and transformation of a host cell, and any such construct as may be desirable to provide embodiments of this invention.

This invention also provides a promoter sequence derived from cell's genome, wherein the promoter sequence corresponds to the regulatory region of a gene that is differentially expressed in the cell as compared to a control cell. The promoters are identified and characterized by: 1) probing a cDNA library with a probe corresponding to the SAGE tag sequence or generating a portion of the desired cDNA by conducting anchored PCR using primers based on the SAGE tag sequence. Examples of cell types wherein differential

expression of a gene is related to promoter function include using the partial cDNA product obtained in step one above as a probe, cloning the extreme 5' end of the cDNA, and also by using the 5' end of the cDNA as a probe, cloning from a genomic library the promoter of the gene that encodes the cDNA. These promoters are identified using the methods described below in combination with standard molecular techniques. Functionally equivalent sequences, as defined above, are further provided by this invention.

In one aspect, the promoter is a sequence derived from the genome of a metastatic cell's genome, wherein the promoter region corresponds to the regulatory region of a gene that is differentially expressed in the cell as compared to the non-metastatic cell. Alternatively, the promoter is a sequence derived from the genome of a non-metastatic cell's genome, wherein the promoter region corresponds to the regulatory region of a gene that is differentially expressed in the cell as compared to the metastatic cell. Table 1 and 2, below are examples of such a sort.

The promoters identified above can be operatively linked to a foreign polynucleotide to compel differential expression of the foreign polynucleotide. A foreign polynucleotide is intended to include any sequence which encodes in whole or in part a polypeptide or protein. It also includes sequences encoding ribozymes and antisense molecules.

Foreign polynucleotides also include therapeutic genes that encode dominant inhibitory oligonucleotides and peptides as well as genes that encode regulatory proteins and oligonucleotides. Generally, gene therapy will involve the transfer of a single therapeutic gene although more than one gene may be necessary for the treatment of particular diseases. In one embodiment, the therapeutic gene is a dominant inhibiting mutant of the wild-type immunosuppressive agent. Alternatively, the therapeutic gene could be a wild-type copy of a defective gene or a functional homolog.

In one aspect, a tag identified by any of Seq. ID Nos. 1 through 5911 corresponds to or comprises a polynucleotide that encodes a polypeptide or protein that is biologically active as an antigen, e.g., a native antigen, an altered antigen, a self-antigen or a tumor-associated antigen. Antigens are

identified by noting the overexpression or cell-specific expression of a tag identified herein. Using the methods described below, the gene comprising or corresponding to the tag is identified, cloned and inserted into an APC. The tag corresponds to an antigen if a CTL response is raised under appropriate experimental conditions. The peptide is confirmed immunogeneic if an appropriate immune response is elecited.

The invention also encompasses co-administration of an immunostimulatory factor and a foreign polynucleotide, both under the control of promoters. In one embodiment, the promoter is an APC specific promoter. In alternative embodiment, the promoters are specific to tissue identified in Tables 1 and 2. The immunostimulatory factors of this invention include any polypeptide factors that modulate immune responses mediated by APC and corresponding T cells. For example, co-stimulatory factors that are differentially expressed in APCs can be used directly to boost the APC functions *in vivo*. Co-stimulatory factors have been described above.

The polynucleotides of the invention can be introduced and expressed in a suitable host cell for generating a cell-based vaccine. These methods are described in more detail below.

The polynucleotides can be conjugated to a detectable marker, e.g., an enzymatic label or a radioisotope for detection of nucleic acid and/or expression of the gene in a cell. A wide variety of appropriate detectable markers are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

The polynucleotides embodied in this invention can be obtained using chemical synthesis, recombinant cloning methods, PCR, or any combination

thereof. Methods of chemical polynucleotide synthesis are well known in the art and need not be described in detail herein. One of skill in the art can use the sequence data provided herein to obtain a desired polynucleotide by employing a DNA synthesizer or ordering from a commercial service.

Polynucleotides comprising a desired sequence can be inserted into a suitable vector, and the vector in turn can be introduced into a suitable host cell for replication and amplification. Polynucleotides can be introduced into host cells by any means known in the art. Cells are transformed by introducing an exogenous polynucleotide by direct uptake, endocytosis, transfection, f-mating or electroporation. Once introduced, the exogenous polynucleotide can be maintained within the cell as a non-integrated vector (such as a plasmid) or integrated into the host cell genome. Amplified DNA can be isolated from the host cell by standard methods. See, e.g., Sambrook, et al. (1989) *supra*. RNA can also be obtained from transformed host cell, or it can be obtained directly from the DNA by using a DNA-dependent RNA polymerase.

The present invention further encompasses a variety of gene delivery vehicles comprising the polynucleotide of the present invention. Gene delivery vehicles include both viral and non-viral vectors such as naked plasmid DNA or DNA/liposome complexes. Vectors are generally categorized into cloning and expression vectors. Cloning vectors are useful for obtaining replicate copies of the polynucleotides they contain, or as a means of storing the polynucleotides in a depository for future recovery. Expression vectors (and host cells containing these expression vectors) can be used to obtain polypeptides produced from the polynucleotides they contain. Suitable cloning and expression vectors include any known in the art, e.g., those for use in bacterial, mammalian, yeast and insect expression systems. The polypeptides produced in the various expression systems are also within the scope of the invention and are described above.

When the vectors are used for gene therapy *in vivo* or *ex vivo*, a pharmaceutically acceptable vector is preferred, such as a replication-incompetent retroviral or adenoviral vector. Pharmaceutically acceptable vectors containing the nucleic acids of this invention can be further modified

for transient or stable expression of the inserted polynucleotide. As used herein, the term "pharmaceutically acceptable vector" includes, but is not limited to, a vector or delivery vehicle having the ability to selectively target and introduce the nucleic acid into dividing cells. An example of such a vector is a "replication-incompetent" vector defined by its inability to produce viral proteins, precluding spread of the vector in the infected host cell. An example of a replication-incompetent retroviral vector is LNL6 (Miller A.D. et al. (1989) *BioTechniques* 7:980-990). The methodology of using replication-incompetent retroviruses for retroviral-mediated gene transfer of gene markers is well established (Correll et al. (1989) *PNAS USA* **86**:8912; Bordignon (1989) *PNAS USA* **86**:8912-52; Culver K. (1991) *PNAS USA* **88**:3155; and Rill, D.R. (1991) *Blood* **79(10)**:2694. Clinical investigations have shown that there are few or no adverse effects associated with the viral vectors, see Anderson (1992) *Science* **256**:808-13.

Compositions containing the polynucleotides of this invention, in isolated form or contained within a vector or host cell are further provided herein. When these compositions are to be used pharmaceutically, they are combined with a pharmaceutically acceptable carrier.

A vector of this invention can contain one or more polynucleotides comprising a sequence selected from SEQ ID NOS. 1 to 5911. It can also contain polynucleotide sequences encoding other polypeptides that enhance, facilitate, or modulate the desired result, such as fusion components that facilitate protein purification, and sequences that increase immunogenicity of the resultant protein or polypeptide.

Also embodied in the present invention are host cells transformed with the vectors as described above. Both prokaryotic and eukaryotic host cells may be used. Prokaryotic hosts include bacterial cells, for example *E. coli* and *Mycobacteria*. Among eukaryotic hosts are yeast, insect, avian, plant and mammalian cells. Host systems are known in the art and need not be described in detail herein. Examples of mammalian host cells include but not limited to COS, HeLa, and CHO cells.

The host cells of this invention can be used, inter alia, as repositories of polynucleotides differentially expressed in non-metastatic or metastatic breast tumor cells, or as vehicles for production of the polynucleotides and the encoded polypeptides.

Methods of Cloning Novel Transcripts and Genes

As noted above, this invention encompasses genes, either genomic or cDNA, which code for a polypeptide or protein in the cell of interest. The genes specifically hybridize under moderate or stringent conditions to a polynucleotide identified by SEQ ID NOS. 1 through 5911 or their respective complements. The process of identification of larger fragment or the full-length coding sequence to which the partial sequence depicted in SEQ ID NOS. 1 through 5911 hybridizes preferably involves the use of the methods and reagents provided in this invention, either singularly or in combination. The complete coding sequence for the gene (either genomic or cDNA) may be known or novel.

RACE-PCR Technique

One method to isolate the gene or cDNA which codes for a polypeptide or protein involves the 5'-RACE-PCR technique. In this technique, the poly-A mRNA that contains the coding sequence of particular interest is first identified by hybridization to a sequence disclosed herein and then reverse transcribed with a 3'-primer comprising the sequence disclosed herein. The newly synthesized cDNA strand is then tagged with an anchor primer of a known sequence, which preferably contains a convenient cloning restriction site attached at the 5'end. The tagged cDNA is then amplified with the 3'-primer (or a nested primer sharing sequence homology to the internal sequences of the coding region) and the 5'-anchor primer. The amplification may be conducted under conditions of various levels of stringency to optimize the amplification specificity. 5'-RACE-PCR can be readily performed using commercial kits (available from, e.g., BRL Life Technologies Inc, Clontech) according to the manufacturer's instructions.

Isolation of partial cDNA (3' fragment) by 3' directed PCR reaction

This procedure is a modification of the protocol described in Polyak et al. (1997) *Nature* **389**:300. Briefly, the procedure uses SAGE tags in PCR reaction such that the resultant PCR product contains the SAGE tag of interest as well as additional cDNA, the length of which is defined by the position of the tag with respect to the 3' end of the cDNA. The cDNA product derived from such a transcript driven PCR reaction can be used for many applications.

RNA from a source believed to express the cDNA corresponding to a given tag is first converted to double-stranded cDNA using any standard cDNA protocol. Similar conditions used to generate cDNA for SAGE library construction can be employed except that a modified oligo-dT primer is used to derive the first strand synthesis. For example, the oligonucleotide of composition 5'-**Biotin**-TCC GGC GCG CCG TTT T CC CAG TCA CGA(30)-3', contains a poly-T stretch at the 3' end for hybridization and priming from poly-A tails, an M13 priming site for use in subsequent PCR steps, a 5' Biotin label (**B**) for capture to strepavidin-coated magnetic beads, and an AscI restriction endonuclease site for releasing the cDNA from the streptavidin-coated magnetic beads. Theoretically, any sufficiently-sized DNA region capable of hybridizing to a PCR primer can be used as well as any other 8 base pair recognizing endonuclease.

cDNA constructed utilizing this or similar modified oligo-dT primer is then processed exactly as described in U.S. Patent No. 5,695,937 up until adapter ligation where only one adapter is ligated to the cDNA pool. After adapter ligation, the cDNA is released from the streptavidin-coated magnetic beads and is then used as a template for cDNA amplification.

Various PCR protocols can be employed using PCR priming sites within the 3' modified oligo-dT primer and the SAGE tag. The SAGE tag-derived PCR primer employed can be of varying length dictated by 5' extension of the tag into the adaptor sequence. cDNA products are now available for a variety of applications.

This technique can be further modified by: (1) altering the length and/or content of the modified oligo-dT primer; (2) ligating adaptors other than that previously employed within the SAGE protocol; (3) performing PCR from template retained on the streptavidin-coated magnetic beads; and (4) priming first strand cDNA synthesis with non-oligo-dT based primers.

Isolation of cDNA using GeneTrapper or modified GeneTrapper Technology

The reagents and manufacturer's instructions for this technology are commercially available from Life Technologies, Inc., Gaithersburg, Maryland. Briefly, a complex population of single-stranded phagemid DNA containing directional cDNA inserts is enriched for the target sequence by hybridization in solution to a biotinylated oligonucleotide probe complementary to the target sequence. The target sequence is based on the tag sequence of the present invention. The hybrids are captured on streptavidin-coated paramagnetic beads. A magnet retrieves the paramagnetic beads from the solution, leaving nonhybridized single-stranded DNAs behind. Subsequently, the captured single-stranded DNA target is released from the biotinylated oligonucleotide. After release, the cDNA clone is further enriched by using a nonbiotinylated target oligonucleotide to specifically prime conversion of the single-stranded target to double-stranded DNA. Following transformation and plating, typically 20% to 100% of the colonies represent the cDNA clone of interest. To identify the desired cDNA clone, the colonies may be screened by colony hybridization using the ^{32}P-labeled oligonucleotide as described above for solution hybridization, or alternatively by DNA sequencing and alignment of all sequences obtained from numerous clones to determine a consensus sequence.

Isolation of cDNAs from a library by probing with the SAGE transcript or tag

Classical methods of constructing cDNA libraries are taught in Sambrook et al., *supra*. Recent procedures described in Velculescu et al. (1997) *Science* **270**:484) can be employed to construct an expression cDNA library cloned into the ZAP Express vector. A ZAP Express cDNA synthesis kit is available from Stratagene is used accordingly to the manufacturer's

protocol. Plates containing 250 to 2000 plaques are hybridized as described in Rupert et al. (1988) *Mol. Cell. Bio.* **8**:3104 to oligonucleotide probes with the same conditions previously described for standard probes except that the hybridization temperature is reduced to room temperature. Washes are performed in 6X standard-saline-citrate 0.1% SDS for 30 minutes at room temperature. The probes are labeled with ^{32}P-ATP through use of T4 polynucleotide kinase.

Identification of known genes or ESTs

In addition, databases exist that reduce the complexity of ESTs by assembling contiguous EST sequences into tentative genes. For example, TIGR has assembled human ESTs into a datable called THC for tentative human consensus sequences. The THC database allows for a more definitive assignment compared to ESTs alone. Software programs exist (TIGR assembler and TIGEM EST assembly machine and contig assembly program (see Huang X. (1996) *Genomics* **33**:21-23)) that allow for assembling ESTs into contiguous sequences from any organism.

Polypeptides of the Invention

This invention provides proteins or polypeptides expressed from a polynucleotide of this invention, which is intended to include wild-type and recombinantly produced polypeptides and proteins from procaryotic and eucaryotic host cells, as well as muteins, analogs, fusions and fragments thereof. In some embodiments, the term also includes antibodies and anti-idiotypic antibodies.

It is understood that equivalents or variants of the wild-type polypeptide or protein also are within the scope of this invention. An "equivalent" varies from the wild-type sequence encoded by the polynucleotides of the invention by any combination of additions, deletions, or substitutions while preserving at least one functional property of the fragment relevant to the context in which it is being used. For instance, an equivalent of a polypeptide of the invention may have the ability to elicit an immune

response with a similar antigen specificity as that elicited by the wild-type polypeptide. As is apparent to one skilled in the art, the equivalent may also be associated with, or conjugated with, other substances or agents to facilitate, enhance, or modulate its function.

The invention includes modified polypeptides containing conservative or non-conservative substitutions that do not significantly affect their properties, such as the immunogenicity of the peptides or their tertiary structures. Modification of polypeptides is routine practice in the art. Amino acid residues which can be conservatively substituted for one another include but are not limited to: glycine/alanine; valine/isoleucine/leucine; asparagine/glutamine; aspartic acid/glutamic acid; serine/threonine; lysine/arginine; and phenylalanine/tryosine. These polypeptides also include glycosylated and nonglycosylated polypeptides, as well as polypeptides with other post-translational modifications, such as, for example, glycosylation with different sugars, acetylation, and phosphorylation.

The polypeptides of the invention can also be conjugated to a chemically functional moiety. Typically, the moiety is a label capable of producing a detectable signal. These conjugated polypeptides are useful, for example, in detection systems such as imaging of breast tumor. Such labels are known in the art and include, but are not limited to, radioisotopes, enzymes, fluorescent compounds, chemiluminescent compounds, bioluminescent compounds substrate cofactors and inhibitors. See, for examples of patents teaching the use of such labels, U.S. Patent Nos. 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149; and 4,366,241. The moieties can be covalently linked to the polypeptides, recombinantly linked, or conjugated to the polypeptides through a secondary reagent, such as a second antibody, protein A, or a biotin-avidin complex.

Other functional moieties include agents that enhance immunological reactivity, agents that facilitate coupling to a solid support, vaccine carriers, bioresponse modifiers, paramagnetic labels and drugs. Agents that enhance immunological reactivity include, but are not limited to, bacterial superantigens. Agents that facilitate coupling to a solid support include, but

are not limited to, biotin or avidin. Immunogen carriers include, but are not limited to, any physiologically acceptable buffers.

The invention also encompasses fusion proteins comprising polypeptides encoded by the polynucleotides disclosed herein and fragments thereof. Such fusion may be between two or more polypeptides of the invention and a related or unrelated polypeptide. Useful fusion partners include sequences that facilitate the intracellular localization of the polypeptide, or enhance immunological reactivity or the coupling of the polypeptide to an immunoassay support or a vaccine carrier. For instance, the polypeptides can be fused with a bioresponse modifier. Examples of bioresponse modifiers include, but are not limited to, fluorescent proteins such as green fluorescent protein (GFP), cytokines or lymphokines such as interleukin-2 (IL-2), interleukin 4 (IL-4), GM-CSF, and K-interferon. Another useful fusion sequence is one that facilitates purification. Examples of such sequences are known in the art and include those encoding epitopes such as Myc, HA (derived from influenza virus hemagglutinin), His-6, or FLAG. Other fusion sequences that facilitate purification are derived from proteins such as glutathione S-transferase (GST), maltose-binding protein (MBP), or the Fc portion of immunoglobulin. For immunological purposes, tandemly repeated polypeptide segments may be used as antigens, thereby producing highly immunogenic proteins.

The proteins of this invention also can be combined with various liquid phase carriers, such as sterile or aqueous solutions, pharmaceutically acceptable carriers, suspensions and emulsions. Examples of non-aqueous solvents include propyl ethylene glycol, polyethylene glycol and vegetable oils. When used to prepare antibodies, the carriers also can include an adjuvant that is useful to non-specifically augment a specific immune response. A skilled artisan can easily determine whether an adjuvant is required and select one. However, for the purpose of illustration only, suitable adjuvants include, but are not limited to Freund's Complete and Incomplete, mineral salts and polynucleotides.

The proteins and polypeptides of this invention are obtainable by a number of processes well known to those of skill in the art, which include purification, chemical synthesis and recombinant methods. Full-length proteins can be purified from a cell derived from non-metastatic or metastatic breast tumor tissue or tissue lysate by methods such as immunoprecipitation with antibody, and standard techniques such as gel filtration, ion-exchange, reversed-phase, and affinity chromatography using a fusion protein as shown herein. For such methodology, see for example Deutscher et al. (1999) GUIDE TO PROTEIN PURIFICATION: METHODS IN ENZYMOLOGY (Vol. 182, Academic Press). Accordingly, this invention also provides the processes for obtaining these proteins and polypeptides as well as the products obtainable and obtained by these processes.

The proteins and polypeptides also can be obtained by chemical synthesis using a commercially available automated peptide synthesizer such as those manufactured by Perkin Elmer/Applied Biosystems, Inc., Model 430A or 431A, Foster City, CA, USA. The synthesized protein or polypeptide can be precipitated and further purified, for example by high performance liquid chromatography (HPLC). Accordingly, this invention also provides a process for chemically synthesizing the proteins of this invention by providing the sequence of the protein and reagents, such as amino acids and enzymes and linking together the amino acids in the proper orientation and linear sequence.

Alternatively, the proteins and polypeptides can be obtained by well-known recombinant methods as described, for example, in Sambrook et al. (1989) *supra*, using the host cell and vector systems described above.

Antibodies

Also provided by this invention is an antibody capable of specifically binding to the proteins or polypeptides as described above. The antibodies of the present invention encompass polyclonal antibodies and monoclonal antibodies. They include but are not limited to mouse, rat, and rabbit or human antibodies. This invention also encompasses functionally equivalent antibodies and fragments thereof. As used herein with respect to the

exemplified antibodies, the phrase "functional equivalent" means a antibody or fragment thereof, or any molecule having the antigen binding site (or epitope) of the antibody that cross-blocks an exemplified antibody when used in an immunoassay such as immunoblotting or immunoprecipitation.

Antibody fragments include the Fab, Fab', $F(ab')_2$, and Fv regions, or derivatives or combinations thereof. Fab, Fab', and $F(ab')_2$ regions of an immunoglobulin may be generated by enzymatic digestion of the monoclonal antibodies using techniques well known to those skilled in the art. Fab fragments may be generated by digesting the monoclonal antibody with papain and contacting the digest with a reducing agent to reductively cleave disulfide bonds. Fab' fragments may be obtained by digesting the antibody with pepsin and reductive cleavage of the fragment so produce with a reducing agent. In the absence of reductive cleavage, enzymatic digestion of the monoclonal with pepsin produces $F(ab')_2$ fragments.

It will further be appreciated that encompassed within the definition of antibody fragment is single chain antibody that can be generated as described in U.S. Pat. No. 4,704,692, as well as chimeric antibodies and humanized antibodies (Oi et al. (1986) *BioTechniques* **4(3)**:214). Chimeric antibodies are those in which the various domains of the antibodies' heavy and light chains are coded for by DNA from more than one species.

As used herein with regard to the monoclonal antibody, the "hybridoma cell line" is intended to include all derivatives, progeny cells of the parent hybridoma that produce the monoclonal antibodies specific for the polypeptides of the present invention, regardless of generation of karyotypic identity.

Laboratory methods for producing polyclonal antibodies and monoclonal antibodies, as well as deducing their corresponding nucleic acid sequences, are known in the art, see Harlow and Lane (1988) *supra* and Sambrook et al. (1989) *supra*. For production of polyclonal antibodies, an appropriate host animal is selected, typically a mouse or rabbit. The substantially purified antigen, whether the whole transmembrane domain, a fragment thereof, or a polypeptide corresponding to a segment of or the entire

specific loop region within the transmembrane domain, coupled or fused to another polypeptide, is presented to the immune system of the host by methods appropriate for the host. The antigen is introduced commonly by injection into the host footpads, via intramuscular, intraperitoneal, or intradermal routes. Peptide fragments suitable for raising antibodies may be prepared by chemical synthesis, and are commonly coupled to a carrier molecule (e.g., keyhole limpet hemocyanin) and injected into a host over a period of time suitable for the production of antibodies. Alternatively, the antigen can be generated recombinantly as a fusion protein. Examples of components for these fusion proteins include, but are not limited to myc, HA, FLAG, His-6, glutathione S-transferease, maltose binding protein or the Fc portion of immunoglobulin.

The monoclonal antibodies of this invention refer to antibody compositions having a homogeneous antibody population. It is not intended to be limited as regards to the source of the antibody or the manner in which it is made. Generally, monoclonal antibodies are biologically produced by introducing protein or a fragment thereof into a suitable host, e.g., a mouse. After the appropriate period of time, the spleens of such animal is excised and individual spleen cells fused, typically, to immortalized myeloma cells under appropriate selection conditions. Thereafter the cells are clonally separated and the supernatants of each clone are tested for their production of an appropriate antibody specific for the desired region of the antigen using methods well known in the art.

The isolation of other hybridomas secreting monoclonal antibodies with the specificity of the monoclonal antibodies of the invention can also be accomplished by one of ordinary skill in the art by producing anti-idiotypic antibodies (Herlyn et al. (1986) *Science* **232**:100). An anti-idiotypic antibody is an antibody which recognizes unique determinants present on the monoclonal antibody produced by the hybridoma of interest.

Idiotypic identity between monoclonal antibodies of two hybridomas demonstrates that the two monoclonal antibodies are the same with respect to their recognition of the same epitopic determinant. Thus, by using antibodies to the epitopic determinants on a monoclonal antibody it is possible to identify

other hybridomas expressing monoclonal antibodies of the same epitopic specificity.

It is also possible to use the anti-idiotype technology to produce monoclonal antibodies which mimic an epitope. For example, an anti-idiotypic monoclonal antibody made to a first monoclonal antibody will have a binding domain in the hypervariable region which is the mirror image of the epitope bound by the first monoclonal antibody. Thus, in this instance, the anti-idiotypic monoclonal antibody could be used for immunization for production of these antibodies.

Other suitable techniques of antibody production include, but are not limited to, *in vitro* exposure of lymphocytes to the antigenic polypeptides or selection of libraries of antibodies in phage or similar vectors. See Huse et al. (1989) *Science* **246**:1275-1281. Genetically engineered variants of the antibody can be produced by obtaining a polynucleotide encoding the antibody, and applying the general methods of molecular biology to introduce mutations and translate the variant. The above described antibody "derivatives" are further provided herein.

Sera harvested from the immunized animals provide a source of polyclonal antibodies. Detailed procedures for purifying specific antibody activity from a source material are known within the art. Undesired activity cross-reacting with other antigens, if present, can be removed, for example, by running the preparation over adsorbants made of those antigens attached to a solid phase and eluting or releasing the desired antibodies off the antigens. If desired, the specific antibody activity can be further purified by such techniques as protein A chromatography, ammonium sulfate precipitation, ion exchange chromatography, high-performance liquid chromatography and immunoaffinity chromatography on a column of the immunizing polypeptide coupled to a solid support.

The specificity of an antibody refers to the ability of the antibody to distinguish polypeptides comprising the immunizing epitope from other polypeptides. An ordinary skill in the art can readily determine without undue experimentation whether an antibody shares the same specificity as a antibody

of this invention by determining whether the antibody being tested prevents an antibody of this invention from binding the polypeptide(s) with which the antibody is normally reactive. If the antibody being tested competes with the antibody of the invention as shown by a decrease in binding by the antibody of this invention, then it is likely that the two antibodies bind to the same or a closely related epitope. Alternatively, one can pre-incubate the antibody of this invention with the polypeptide(s) with which it is normally reactive, and determine if the antibody being tested is inhibited in its ability to bind the antigen. If the antibody being tested is inhibited, then, in all likelihood, it has the same, or a closely related, epitopic specificity as the antibody of this invention.

The antibodies of the invention can be bound to many different carriers. Thus, this invention also provides compositions containing antibodies and a carrier. Carriers can be active and/or inert. Examples of well-known carriers include polypropylene, polystyrene, polyethylene, dextran, nylon, amylases, glass, natural and modified celluloses, polyacrylamides, agaroses and magnetite. The nature of the carrier can be either soluble or insoluble for purposes of the invention. Those skilled in the art will know of other suitable carriers for binding antibodies, or will be able to ascertain such, using routine experimentation.

The antibodies of this invention can also be conjugated to a detectable agent or a hapten. The complex is useful to detect the polypeptide(s) (or polypeptide fragments) to which the antibody specifically binds in a sample, using standard immunochemical techniques such as immunohistochemistry as described by Harlow and Lane (1988). *supra*. There are many different labels and methods of labeling known to those of ordinary skill in the art. Examples of the types of labels which can be used in the present invention include radioisotopes, enzymes, colloidal metals, fluorescent compounds, bioluminescent compounds, and chemiluminescent compounds. Those of ordinary skill in the art will know of other suitable labels for binding to the antibody, or will be able to ascertain such, using routine experimentation.

Furthermore, the binding of these labels to the antibody of the invention can be done using standard techniques common to those of ordinary skill in the art.

Another technique which may also result in greater sensitivity consists of coupling the antibodies to low molecular weight haptens. These haptens can then be specifically detected by means of a second reaction. For example, it is common to use such haptens as biotin, which reacts avidin, or dinitropherryl, pyridoxal, and fluorescein, which can react with specific anti-hapten antibodies. See Harlow and Lane (1988) *supra*.

Compositions containing the antibodies, fragments thereof or cell lines which produce the antibodies, are encompassed by this invention. When these compositions are to be used pharmaceutically, they are combined with a pharmaceutically acceptable carrier.

Uses of polynucleotides, polypeptides and antibodies of the present invention

The polynucleotides, polypeptides and antibodies embodied in this invention provide specific reagents that can be used in standard diagnostic procedures. Accordingly, one embodiment of the present invention is a method of diagnosing the metastatic condition of a breast cell by detecting differential expression of a polynucleotide comprising any one of the sequences listed in SEQ ID NOS. 1 to 5911, or 1-3175 or 3176-5911, or the populations identified above, or the encoded polypeptide(s). The method can be used for aiding in the diagnosis of metastatic breast cancer by detecting a genotype that is correlated with a phenotype characteristic of metastatic breast tumor cells.

In one aspect, overexpression of a polynucleotide identified in Table 2 or comprising or corresponding to Seq. ID No. 3176-5911 is indicative of the non-metastatic state of a breast cell. Conversely, overexpression of a polynucleotide comprising the sequence selected from polynucleotide (e.g., identified in Table 1 or comprising or corresponding to Seq. ID No. 1 to 3175) is indicative of the non-metastatic state of a breast cell.

In yet another aspect, the differential expression of the polynucleotides is determined by assaying for a difference, between the non-metastatic and metastatic breast tumor cells, in the level of transcripts that specifically hybridize with one or more of the exemplified sequences. In another aspect, the differential expression of the polynucleotides is determined by detecting a difference in the level of the encoded polypeptides.

Cell or tissue samples used for this invention encompass body fluid, solid tissue samples, tissue cultures or cells derived therefrom and the progeny thereof, and sections or smears prepared from any of these sources, or any other samples that may contain a breast cell having the polynucleotides disclosed herein or their gene products.

In assaying for an alteration in mRNA level, nucleic acid contained in the aforementioned samples is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), *supra* or extracted by nucleic-acid-binding resins following the accompanying instructions provided by manufactures. The mRNA contained in the extracted nucleic acid sample is then detected by hybridization (e.g. Northern blot analysis) and/or amplification procedures according to methods widely known in the art or based on the methods exemplified herein.

Nucleic acid molecules having at least 10 nucleotides and exhibiting sequence complementarity or homology to the polynucleotides described herein find utility as hybridization probes. It is known in the art that a "perfectly matched" probe is not needed for a specific hybridization. Minor changes in probe sequence achieved by substitution, deletion or insertion of a small number of bases do not affect the hybridization specificity. In general, as much as 20% base-pair mismatch (when optimally aligned) can be tolerated. Preferably, a probe useful for detecting the aforementioned mRNA that is differentially expressed in non-metastatic or metastatic breast tissues is at least about 80% identical to the homologous region of comparable size contained in the sequences to be detected. More preferably, the probe is 85% identical to the corresponding gene sequence after alignment of the homologous region;

even more preferably, it exhibits 90% identity. Specifically, a preferred probe is selected from the group of SEQ ID NOS. 1 to 5911, or their respective complements.

These probes can be used in hybridization reaction (*e.g.* Southern and Northern blot analysis) to detect, prognose, diagnose or monitor the metastatic states associated with the differential expression of these genes. The total size of fragment, as well as the size of the complementary stretches, will depend on the intended use or application of the particular nucleic acid segment. Smaller fragments derived from the known sequences will generally find use in hybridization embodiments, wherein the length of the complementary region may be varied, such as between about 10 and about 100 nucleotides, or even full length according to the complementary sequences one wishes to detect.

Nucleotide probes having complementary sequences over stretches greater than 10 nucleotides in length are generally preferred, so as to increase stability and selectivity of the hybrid, and thereby improving the specificity of particular hybrid molecules obtained. More preferably, one can design nucleic acid molecules having gene-complementary stretches of more than 50 nucleotides in length, or even longer where desired. Such fragments may be readily prepared by, for example, directly synthesizing the fragment by chemical means, by application of nucleic acid reproduction technology, such as the PCR™ technology with two priming oligonucleotides as described in U.S. Pat. No. 4,603,102 or by introducing selected sequences into recombinant vectors for recombinant production. A preferred probe is about 50-75 or more preferably, 50-100, nucleotides in length.

In certain embodiments, it will be advantageous to employ nucleic acid sequences of the present invention in combination with an appropriate means, such as a label, for detecting hybridization and therefore complementary sequences. A wide variety of appropriate indicator means are known in the art, including fluorescent, radioactive, enzymatic or other ligands, such as avidin/biotin, which are capable of giving a detectable signal. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as urease, alkaline phosphatase or peroxidase, instead of

radioactive or other environmental undesirable reagents. In the case of enzyme tags, colorimetric indicator substrates are known which can be employed to provide a means visible to the human eye or spectrophotometrically, to identify specific hybridization with complementary nucleic acid-containing samples.

The nucleotide probes of the present invention can also be used as primers and detection of genes or gene transcripts that are differentially expressed in certain body tissues. A preferred primer is one comprising a sequence of SEQ ID NOS. 1 through 5911 or their respective complements. Additionally, a primer useful for detecting the aforementioned gene or transcript is at least about 80% identical to the homologous region of comparable size of the gene or transcript to be detected contained in the previously identified sequences. For the purpose of this invention, amplification means any method employing a primer-dependent polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA-polymerases such as T7 DNA polymerase, Klenow fragment of *E.coli* DNA polymerase, and reverse transcriptase.

A preferred amplification method is PCR. General procedures for PCR are taught in MacPherson et al., PCR: A PRACTICAL APPROACH, (IRL Press at Oxford University Press (1991)). However, PCR conditions used for each application reaction are empirically determined. A number of parameters influence the success of a reaction. Among them are annealing temperature and time, extension time, Mg^{2+} ATP concentration, pH, and the relative concentration of primers, templates, and deoxyribonucleotides.

After amplification, the resulting DNA fragments can be detected by agarose gel electrophoresis followed by visualization with ethidium bromide staining and ultraviolet illumination. A specific amplification of the gene or transcript of interest can be verified by demonstrating that the amplified DNA fragment has the predicted size, exhibits the predicated restriction digestion pattern, and/or hybridizes to the correct cloned DNA sequence.

The probes also can be attached to a solid support for use in high throughput screening assays using methods known in the art. PCT WO 97/10365 and U.S. Patent numbers 5,405,783, 5,412,087 and 5,445,934, for example, disclose the construction of high density oligonucleotide chips which can contain one or more of the sequences disclosed herein. Based in the methods disclosed in U.S. Patent numbers 5,405,783, 5,412,087 and 5,445,934, the probes of this invention are synthesized on a derivatized glass surface. Photoprotected nucleoside phosphoramidites are coupled to the glass surface, selectively deprotected by photolysis through a photolithographic mask, and reacted with a second protected nucleoside phosphoramidite. The coupling/deprotection process is repeated until the desired probe is complete.

The expression level of a gene of interest is determined through exposure of a nucleic acid sample to the probe-modified chip. Extracted nucleic acid is labeled, for example, with a fluorescent tag, preferably during an amplification step. Hybridization of the labeled sample is performed at an appropriate stringency level. The degree of probe-nucleic acid hybridization is quantitatively measured using a detection device, such as a confocal microscope. See U.S. Pat Nos. 5,578,832 and 5,631,734. The obtained measurement is directly correlated with gene expression level.

More specifically, the probes and high density oligonucleotide probe arrays provide an effective means of monitoring expression of a multiplicity of genes. The expression monitoring methods of this invention may be used in a wide variety of circumstances including detection of disease, identification of differential gene expression between two samples, or screening for compositions that upregulate or downregulate the expression of particular genes.

In another preferred embodiment, the methods of this invention are used to monitor expression of the genes which specifically hybridize to the probes of this invention in response to defined stimuli, such as a drug.

In one embodiment, the hybridized nucleic acids are detected by detecting one or more labels attached to the sample nucleic acids. The labels may be incorporated by any of a number of means well known to those of skill

in the art. However, in one aspect, the label is simultaneously incorporated during the amplification step in the preparation of the sample nucleic acid. Thus, for example, polymerase chain reaction (PCR) with labeled primers or labeled nucleotides will provide a labeled amplification product. In a separate embodiment, transcription amplification, as described above, using a labeled nucleotide (e.g. fluorescein-labeled UTP and/or CTP) incorporates a label in to the transcribed nucleic acids.

Alternatively, a label may be added directly to the original nucleic acid sample (e.g., mRNA, polyA, mRNA, cDNA, etc.) or to the amplification product after the amplification is completed. Means of attaching labels to nucleic acids are well known to those of skill in the art and include, for example nick translation or end-labeling (e.g. with a labeled RNA) by kinasing of the nucleic acid and subsequent attachment (ligation) of a nucleic acid linker joining the sample nucleic acid to a label (e.g., a fluorophore).

The nucleic acid sample also may be modified prior to hybridization to the high density probe array in order to reduce sample complexity thereby decreasing background signal and improving sensitivity of the measurement using the methods disclosed in WO 97/10365.

Results from the chip assay are typically analyzed using a computer software program. See, for example, EP 0717 113 A2 and WO 95/20681. The hybridization data are read into the program, which calculates the expression level of the targeted gene(s). This figure is compared against existing data sets of gene expression levels for diseased and healthy individuals. A correlation between the obtained data and that of a set of diseased individuals having non-metastatic or metastatic breast cancer indicates the neoplastic stage of the tested tumor sample.

Expression of the genes associated with breast cancer progression can also be determined by examining the protein product of the polynucleotides of the present invention. Determining the protein level involves a) providing a biological sample containing polypeptides; and (b) measuring the amount of any immunospecific binding that occurs between an antibody reactive to the

protein products of interest and a component in the sample, in which the amount of immunospecific binding indicates the level of the protein products.

A variety of techniques are available in the art for protein analysis. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using *e.g.*, colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunoflourescent assays, and SDS-PAGE. In addition, cell sorting analysis can be employed to detect cell surface antigens. Such analysis involves labeling target cells with antibodies coupled to a detectable agent, and then separating the labeled cells from the unlabeled ones in a cell sorter. A sophisticated cell separation method is fluorescence-activated cell sorting (FACS). Cells traveling in single file in a fine stream are passed through a laser beam, and the fluorescence of each cell bound by the fluorescently labeled antibodies is then measured.

Antibodies that specifically recognize and bind to the protein products of interest are required for conducting the aforementioned protein analyses. These antibodies may be purchased from commercial vendors or generated and screened using methods well known in the art. See Harlow and Lane (1988) *supra.* and Sambrook et al. (1989) *supra.*

In diagnosing malignancy or metastasis characterized by a differential expression of genes or transcripts that are associated with either the non-metastatic or metastatic state of a breast cell, one typically conducts a comparative analysis of the subject and appropriate controls. Preferably, a diagnostic test includes a control sample derived from a subject (hereinafter positive control), that exhibits a detectable increase in expression of the genes, preferably at a level of 3 folds or more and clinical characteristics of tumor metastasis. More preferably, a diagnosis also includes a control sample derived from a subject (hereinafter negative control), that lacks the clinical characteristics of the metastatic state and whose expression level of the gene at question is within a normal range. A positive correlation between the subject and the positive control with respect to the identified differential gene

expression indicates the presence or a predisposition of metastatic breast cancer. A lack of correlation between the subject and the negative control confirms the diagnosis.

The selection of an appropriate control cell or tissue is dependent on the sample cell or tissue initially selected and its phenotype which is under investigation. Whereas the sample cell is derived from a metastatic breast tumor tissue, one or more counterpart non-metastatic cells of the sample cells can be used as control cells. Counterparts would include, for example, cell lines established from the same or related cells to those found in the sample cell population. Preferably, a control matches the tissue, and/or cell type the tested sample is derived from. More preferably, a control is derived from a primary breast tumor of the same individual from whom the test sample is derived. It is also preferable to analyze the control and the tested sample in parallel.

There are various methods available in the art for quantifying mRNA or protein level from a cell sample and indeed, any method that can quantify these levels is encompassed by this invention. For example, determination of the mRNA level of the gene may involve, in one aspect, measuring the amount of mRNA in a mRNA sample isolated from the breast cell by hybridization or quantitative amplification using at least one oligonucleotide probe that is complementary to the mRNA. Determination of the aforementioned protein products requires measuring the amount of immunospecific binding that occurs between an antibody reactive to the product of interest. To detect and quantify the immunospecific binding, or signals generated during hybridization or amplification procedures, digital image analysis systems including but not limited to those that detect radioactivity of the probes or chemiluminescence can be employed.

Screening Assays

The present invention also provides a screen for various agents which modulate the expression of a polynucleotide associated the metastatic condition of a breast cell by first contacting a cell with an effective amount of

a potential agent, and then assaying for a change in the expression level of a polynucleotide selected from the populations identified above. A change in the expression level is indicative of a candidate therapeutic agent. Preferably, the agent when administered into a cell or subject reduces the level of expression of a gene or transcript that is associated with breast cancer progression and is further characterized as comprising a sequence selected from SEQ ID NO. 1 through 3175. A preferred agent may also enhance expression of genes or transcripts comprising a sequence of SEQ ID NOS. 3176 to 5911. In certain aspects of the invention, an agent may result in phenotypic changes of the recipient cell as evidenced by an agent-induced cell apoptosis, a reduced rate of cell growth or cell motility. Altered gene expression can be detected by assaying for altered mRNA expression or protein expression using the probes, primers and antibodies as described herein.

To practice the method *in vitro*, suitable cell cultures or tissue cultures from metastatic breast cells are first provided. The cell can be a cultured cell or a genetically modified cell in which a transcript from SEQ ID NOS. 1 through 5911, or their complements, or alternatively, transcripts which contain or correspond to a tag or its respective complement is expressed. Alternatively, the cells can be from a tissue biopsy. The cells are cultured under conditions (temperature, growth or culture medium and gas (CO_2)) and for an appropriate amount of time to attain exponential proliferation without density dependent constraints. It also is desirable to maintain an additional separate cell culture; one which does not receive the agent being tested as a control.

As is apparent to one of skill in the art, suitable cells may be cultured in microtiter plates and several agents may be assayed at the same time by noting genotypic changes and/or phenotypic changes.

When the agent is a composition other than naked DNA or RNA, the agent may be directly added to the cell culture or added to culture medium for addition. As is apparent to those skilled in the art, an "effective" amount must be added which can be empirically determined. When the agent is a

polynucleotide, it may be introduced directly into a cell by transfection or electroporation. Alternatively, it may be inserted into the cell using a gene delivery vehicle or other methods as described above.

For the purposes of this invention, an "agent" is intended to include, but not be limited to a biological or chemical compound such as a simple or complex organic or inorganic molecule, a peptide, a protein (e.g. antibody) or a polynucleotide (e.g. anti-sense). A vast array of compounds can be synthesized, for example polymers, such as polypeptides and polynucleotides, and synthetic organic compounds based on various core structures, and these are also included in the term "agent". In addition, various natural sources can provide compounds for screening, such as plant or animal extracts, and the like. It should be understood, although not always explicitly stated that the agent is used alone or in combination with another agent, having the same or different biological activity as the agents identified by the inventive screen. The agents and methods also are intended to be combined with other therapies.

The assays also can be performed in a subject. When the subject is an animal such as a rat, mouse or simian, the method provides a convenient animal model system which can be used prior to clinical testing of an agent. In this system, a candidate agent is a potential drug if transcript expression is altered, i.e., upregulated (such as restoring tumor suppressor function), downregulated or eliminated as with drug resistant genes or oncogenes, or if symptoms associated or correlated to the presence of cells containing transcript expression are ameliorated, each as compared to untreated, animal having the pathological cells. It also can be useful to have a separate negative control group of cells or animals which are healthy and not treated, which provides a basis for comparison. After administration of the agent to subject, suitable cells or tissue samples are collected and assayed for altered gene expression.

As an example of an animal model, groups of nude mice (Balb/c NCR nu/nu female, Simonsen, Gilroy, CA) are each subcutaneously inoculated with about 10^5 to about 10^9 hyperproliferative, cancer or target cells as defined herein. When the tumor is established, the agent is administered, for example, by subcutaneous injection around the tumor. Tumor measurements to

determine reduction of tumor size are made in two dimensions using venier calipers twice a week. Other animal models may also be employed as appropriate.

These agents of this invention and the above noted compounds and their derivatives can be combined with a pharmaceutically acceptable carrier for the preparation of medicaments for use in the methods described herein. They can be administered to treat a cancerous condition, or to prevent progression from a pre-neoplastic or non-metastatic state into a neoplastic or a metastatic state.

In a preferred embodiment, an agent of the present invention is administered to reverse the metastatic condition of a breast cell. As used herein, the term "reversing the metastatic condition" of a cell is intended to include apoptosis, necrosis or any other means of preventing cell division, reduced cell motility, loss of pharmaceutical resistance, maturation, differentiation or reversion of any other metastatic phenotypes. For example, characteristics associated with a metastatic phenotype (a set of *in vitro* characteristics associated with a tumorigenic ability *in vivo*) include but are not limited to a more rounded cell morphology, looser substratum attachment, loss of contact inhibition, and loss of anchorage dependence.

One can determine if reversion of the metastatic condition of a breast cell is achieved by performing assays standard in the art. For example, cell proliferation can be assayed by measuring ^{3}H-thymidine incorporation, by direct cell count, by detecting changes in transcriptional activity of known genes such as proto-oncogenes (e.g., fos, myc) or cell cycle markers; cell viability can be assessed by staining cells with a dye that reacts with either living or dead cells; cellular differentiation can be monitored by histological methods or by detecting the presence or loss of certain surface markers that are associated with undifferentiated or differentiated phenotype; cell motility can be assayed directly by measuring the cell migration speed, or indirectly by determining the fraction of cells developed lamellipodia.

The agents of the present invention can be administered to a cell or a subject by various delivery systems known in the art. Non-limiting examples

include encapsulation in liposomes, microparticles, microcapsules, expression by recombinant cells, receptor-mediated endocytosis (see, *e.g.*, Wu and Wu (1987) *J. Biol. Chem.* **262**:4429-4432), and construction of a therapeutic nucleic acid as part of a retroviral or other vector. Methods of delivery include but are not limited to transdermally, gene therapy, intra-arterial, intra-muscular, intravenous, intranasal, and oral routes, and include sustained delivery systems. In a specific embodiment, it may be desirable to administer the pharmaceutical compositions of the invention locally to the area in need of treatment; this may be achieved by, for example, and not by way of limitation, local infusion during surgery, by injection, or by means of a catheter or targeted gene delivery of the sequence coding for the therapeutic.

The agents identified herein as effective for their intended purpose can be administered to subjects or individuals susceptible to or at risk of developing breast cancer. When the agent is administered to a subject such as a mouse, a rat or a human patient, the agent can be added to a pharmaceutically acceptable carrier and systemically or topically administered to the subject. Therapeutic amounts can be empirically determined and will vary with the pathology being treated, the subject being treated and the efficacy and toxicity of the agent.

Administration *in vivo* can be effected in one dose, continuously or intermittently throughout the course of treatment. Methods of determining the most effective means and dosage of administration are well known to those of skill in the art and will vary with the composition used for therapy, the purpose of the therapy, the target cell being treated, and the subject being treated. Single or multiple administrations can be carried out with the dose level and pattern being selected by the treating physician. Suitable dosage formulations and methods of administering the agents can be found below.

The agents and compositions of the present invention can be used in the manufacture of medicaments and for the treatment of humans and other animals by administration in accordance with conventional procedures, such as an active ingredient in pharmaceutical compositions.

The pharmaceutical compositions can be administered orally, intranasally, parenterally, transdermally or by inhalation therapy, and may take the form of tablets, lozenges, granules, capsules, pills, ampoules, suppositories or aerosol form. They may also take the form of gene therapy, suspensions, solutions and emulsions of the active ingredient in aqueous or nonaqueous diluents, syrups, granulates or powders. In addition to an agent of the present invention, the pharmaceutical compositions can also contain other pharmaceutically active compounds or a plurality of compounds of the invention.

It should be understood that in addition to the ingredients particularly mentioned above, the formulations of this invention may include other agents conventional in the art having regard to the type of formulation in question, for example, those suitable for oral administration may include such further agents as sweeteners, thickeners and flavoring agents. It also is intended that the agents, compositions and methods of this invention be combined with other suitable compositions and therapies.

Non-Human Transgenic Animals

In another aspect, the novel polynucleotide sequences associated with non-metastatic and metastatic breast cancer can be used to generate transgenic animal models. In recent years, geneticists have succeeded in creating transgenic animals, for example mice, by manipulating the genes of developing embryos and introducing foreign genes into these embryos. Once these genes have integrated into the genome of the recipient embryo, the resulting embryos or adult animals can be analyzed to determine the function of the gene. The mutant animals are produced to understand the function of known genes *in vivo* and to create animal models of human diseases. (*see, e.g.,* Chisaka *et al.* (1992) **355**:516-520; Joyner et al. (1992) in POSTIMPLANTATION DEVELOPMENT IN THE MOUSE (Chadwick and Marsh, eds., John Wiley & Sons, United Kingdom) pp:277-297; Dorin et al. (1992) Nature **359**:211-215).

Genomics Applications

A cell's transcriptome offers a snapshot of all expressed genes and their relative level of expression. This information provides a library for the study of the number and types of genes whose transcription is induced or regulated during cell processes such as activation, differentiation, aging, viral transformation, morphogenesis, and mitosis. A comparison of the transcriptomes of a particular cell at various times during the life of the cell, under the same or different environmental stimuli, provides insight into the regulatory process of the cell. Using the transcripts provided herein, the analysis of these and other cellular processes and the effects of environmental stimuli on the cell is possible.

This invention also provides a process for preparing a database for the analysis of a cell's expressed genes by storing in a digital storage medium information related to the sequences of the transcriptome. Using this method, a data processing system for standardized representation of the expressed genes of a cell is compiled. The data processing system is useful to analyze gene expression between two cells by first selecting a cell and then identifying and sequencing the transcriptome of the cell. This information is stored in a computer-readable storage medium as the transcriptome. The transcriptome is then compared with at least one sequence(s) of transcription fragments from a reference cell. The compared sequences are then analyzed. Uniquely expressed sequences and sequences differentially expressed between the reference cell and the selected cell can be identified by this method.

In other words, this invention provides a computer based method for screening the homology of an unknown DNA or mRNA sequence against the complete set of expressed genes of a preselected cell by first providing the complete set of expressed genes, i.e., the transcriptome, in computer readable form and homology screening the DNA or mRNA of the unknown sequence against transcriptome and determining whether the DNA sequence of the unknown contains similarities to any portion of the transcriptome listed in the computer readable form.

Thus, the information provided herein also provides a means to compare the relative abundance of gene transcripts in different biological specimens by use of high-throughput sequence-specific analysis of individual RNAs or their corresponding cDNAs using a modification of the systems described in WO 95/2068, 96/23078 and 5,618,672.

The tags or transcripts also can be attached to a solid support for use in high throughput screening assays. PCT WO 97/10365, for example, discloses the construction of high density oligonucleotide chips. See also, U.S. Pat. Nos. 5,405,783, 5,412,087 and 5,445,934. Using this method, the probes are synthesized on a derivatized glass surface. Photoprotected nucleoside phosphoramidites are coupled to the glass surface, selectively deprotected by photolysis through a photolithographic mask, and reacted with a second protected nucleoside phosphoramidite. The coupling/deprotection process is repeated until the desired probe is complete.

The expression level of a gene is determined through exposure of a nucleic acid sample to the probe-modified chip. Extracted nucleic acid is labeled, for example, with a fluorescent tag, preferably during an amplification step. Hybridization of the labeled sample is performed at an appropriate stringency level. The degree of probe-nucleic acid hybridization is quantitatively measured using a detection device, such as a confocal microscope. See U.S. Pat Nos. 5,578,832 and 5,631,734. The obtained measurement is directly correlated with gene expression level.

Results from the chip assay are typically analyzed using a computer software program. See, for example, EP 0717 113 A2 and WO 95/20681. The hybridization data is read into the program, which calculates the expression level of the targeted gene(s). This figure is compared against existing data sets of gene expression levels for that cell type.

For example, the database and methods of using the database provides a means to differentiate normal metastatic from pleural effusion cells from abnormal metastatic from pleural effusion cells. It also allows one to differentiate between metastatic from pleural effusion cells biopsied from different regions from a patient or subject or gene expression before or after

treatment with a potential therapeutic agent. It can be used to analyze drug toxicity and efficacy, as well as to selectively look at protein categories which are expected to be affected by a drug or which may be overexpressed as a result of treatment with a drug, such as the various multi-drug resistant genes. Additional utilities of the database include, but are not limited to analysis of the developmental state of a test cell, the influence of viral or bacterial infection, control of cell cycle, effect of a tumor suppressor gene or lack thereof, polymorphism within the cell type, apoptosis, and the effect of regulatory genes.

Vaccines for Cancer Treatment and Prevention

In one embodiment, the present invention comprises vaccines for cancer treatment. Recent advances in vaccine adjuvants provide effective means of administering peptides so that they impact maximally on the immune system. Del-Giudice (1994) Experientia **50**:1061-1066. A polynucleotide encoding the antigenic peptide also can be administerd as a cancer vaccine. The polynucleotide can be administered as naked DNA or alternatively, in expression vectors. Therapy can be enhanced by coadministration of cytokines and/or co-stimulatory molecules which in turn, can be administered as proteins or the polynucleotides encoding the proteins.

Host Cells comprising Antigenic Peptides of the Invention

The invention further provides isolated host cells comprising antigenic peptides of the invention. In some embodiments, these host cells present one or more peptides of the invention on the surface of the cell in the context of an MHC molecule, i.e., a antigenic peptide of the invention is bound to a cell surface MHC molecule such that the peptide can be recognized by an immune effector cell. Isolated host cells which present the polypeptides of this invention in the context of MHC molecules are further useful to expand and isolate a population of educated, antigen-specific immune effector cells. The immune effector cells, e.g., cytotoxic T lymphocytes, are produced by culturing naïve immune effector cells with antigen-presenting cells cells which

present the polypeptides in the context of MHC molecules on the surface of the APCs. The population can be purified using methods known in the art, e.g., FACS analysis or FICOLL™ gradient. The methods to generate and culture the immune effector cells as well as the populations produced thereby also are the inventors' contribution and invention. Pharmaceutical compositions comprising the cells and pharmaceutically acceptable carriers are useful in adoptive immunotherapy. Prior to administration *in vivo,* the immune effector cells are screened *in vitro* for their ability to lyse melanoma tumor cells.

Gene transfer

Vectors useful in genetic modification

In one embodiment, the present invention provides methods of eliciting efficient antigen-specific immune response in a subject by introducing to the subject recombinant polynucleotides encoding antigenic peptides alone or in combination with immunostimulatory factors. Methods and materials for gene transfer are known in the art, including, for example, viral mediated gene transfer, lipofection, transformation, transfection and transduction. The polynucleotides encoding the immunostimulatory factor and target antigenic peptide can be introduced ex vivo into a host cell, for example, dendritic cells. The genetically modified host cells can be introduced as a cell-based vaccine into the target subject. Alternatively, the polynucleotides encoding the immunostimulatory factor and target antigenic peptidecan be introduced directly into the subject in the form of gene-based vaccine.

Various viral infection techniques have been developed which utilize recombinant viral vectors for gene delivery, and constitute preferred approaches to the present invention. The viral vectors which have been used in gene transfer include, but not limited to, viral sequences derived from simian virus 40 (SV40), adenovirus, adeno-associated virus (AAV), and retroviruses.

Vector Transduction of Cells such as APCs

APCs can be transduced with viral vectors encoding a relevant polypeptides. The most common viral vectors include recombinant poxviruses such as vaccinia and fowlpox virus (Bronte et al. (1997) Proc. Natl. Acad. Sci. USA **94**:3183-3188; Kim et al. (1997) J. Immunother. **20**:276-286) and, preferentially, adenovirus (Arthur et al. (1997) J. Immunol. **159**:1393-1403; Wan et al. (1997) Human Gene Therapy **8**:1355-1363; Huang et al. (1995) J. Virol. **69**:2257-2263). Retrovirus also may be used for transduction of human APCs (Marin et al. (1996) J. Virol. **70**:2957-2962).

In vitro or *ex vivo* exposure of human DCs to adenovirus (Ad) vector at a multiplicity of infection (MOI) of 500 for 16-24 h in a minimal volume of serum-free medium reliably gives rise to foreign polynucleotide expression in 90-100% of DCs. The efficiency of transduction of DCs or other APCs can be assessed by immunofluorescence using fluorescent antibodies specific for the tumor antigen being expressed (Kim et al. (1997) J. Immunother. **20**:276-286). Alternatively, the antibodies can be conjugated to an enzyme (e.g. HRP) giving rise to a colored product upon reaction with the substrate. The actual amount of antigenic polypeptides being expressed by the APCs can be evaluated by ELISA.

In vivo transduction of DCs, or other APCs, can be accomplished by administration of Ad (or other viral vectors) via different routes including intravenous, intramuscular, intranasal, intraperitoneal or cutaneous delivery. The preferred method is cutaneous delivery of Ad vector at multiple sites using a total dose of approximately $1x10^{10}$-$1x\ 10^{12}$ i.u. Levels of *in vivo* transduction can be roughly assessed by co-staining with antibodies directed against APC marker(s) and the antigen being expressed. The staining procedure can be carried out on biopsy samples from the site of administration or on cells from draining lymph nodes or other organs where APCs (in particular DCs) may have migrated (Condon et al. (1996) Nature Med. **2**:1122-1128; Wan et al. (1997) Human Gene Therapy **8**:1355-1363). The amount of antigen being expressed at the site of injection or in other organs where

transduced APCs may have migrated can be evaluated by ELISA on tissue homogenates.

Although viral gene delivery is more efficient, DCs can also be transduced *in vitro/ex vivo* by non-viral gene delivery methods such as electroporation, calcium phosphate precipitation or cationic lipid/plasmid DNA complexes (Arthur et al. (1997) Cancer Gene Therapy **4**:17-25). Transduced APCs can subsequently be administered to the host via an intravenous, subcutaneous, intranasal, intramuscular or intraperitoneal route of delivery.

In vivo transduction of DCs, or other APCs, can potentially be accomplished by administration of cationic lipid/plasmid DNA complexes delivered via the intravenous, intramuscular, intranasal, intraperitoneal or cutaneous route of administration. Gene gun delivery or injection of naked plasmid DNA into the skin also leads to transduction of DCs (Condon et al. (1996) Nature Med. **2**:1122-1128 and Raz et al. (1994) Proc. Natl. Acad. Sci. USA 91:9519-9523). Intramuscular delivery of plasmid DNA may also be used for immunization (Rosato et al. (1997) Human Gene Therapy **8**:1451-1458.

The transduction efficiency and levels of foreign polynucleotide expression can be assessed as described above for viral vectors.

Administration of Cell-Based Vaccine to Subject

Genetically modified cells can subsequently be administered to the host subject via various routes, including, for example, intravenous infusion, subcutaneous injection, intranasal, intramuscular or intraperitoneal delivery. The cells containing the recombinant polynucleotides may be used to confer immunity to individuals. Administration *in vivo* can be effected in one dose, continuously or intermittently throughout the course of treatment. Methods of determining the most effective means and dosage of administration are well known to those of skill in the art and will vary with the composition used for therapy, the purpose of the therapy, the target cell being treated, and the

subject being treated. Single or multiple administrations can be carried out with the dose level and pattern being selected by the treating physician.

Adoptive Immunotherapy Methods

Expanded populations of antigen-specific immune effector cells and APCs presenting antigens find use in adoptive immunotherapy regimes.

Adoptive immunotherapy methods involve, in one aspect, administering to a subject a substantially pure population of educated, antigen-specific immune effector cells made by culturing naïve immune effector cells with APCs as described above. In some embodiments, the APCs are dendritic cells.

In one embodiment, the adoptive immunotherapy methods described herein are autologous. In this case, the APCs are made using parental cells isolated from a single subject. The expanded population also employs T cells isolated from that subject. Finally, the expanded population of antigen-specific cells is administered to the same patient.

In a further embodiment, APCs or immune effector cells are administered with an effective amount of a stimulatory cytokine, such as IL-2 or a co-stimulatory molecule.

Immune Effector Cells

The present invention makes use of antigen-presenting matrices, including APCs, to stimulate production of an enriched population of antigen-specific immune effector cells. Accordingly, the present invention provides a population of cells enriched in educated, antigen-specific immune effector cells, specific for an antigenic peptide of the invention. These cells can cross-react with (bind specifically to) antigenic determinants (epitopes) on natural (endogenous) antigens. In some embodiments, the natural antigen is on the surface of tumor cells and the educated, antigen-specific immune effector cells of the invention suppress growth of the tumor cells. When APCs are used, the antigen-specific immune effector cells are expanded at the expense of the APCs, which die in the culture. The process by which naïve immune effector

cells become educated by other cells is described essentially in Coulie (1997) Molec. Med. Today 3:261-268.

An effector cell population suitable for use in the methods of the present invention can be autogeneic or allogeneic, preferably autogeneic. When effector cells are allogeneic, preferably the cells are depleted of alloreactive cells before use. This can be accomplished by any known means, including, for example, by mixing the allogeneic effector cells and a recipient cell population and incubating them for a suitable time, then depleting $CD69^+$ cells, or inactivating alloreactive cells, or inducing anergy in the alloreactive cell population.

Hybrid immune effector cells can also be used. Immune effector cell hybrids are known in the art and have been described in various publications. See, for example, International Patent Application Nos. WO 98/46785; and WO 95/16775.

The following examples are intended to illustrate, but not limit, the invention.

Examples

SAGE Analysis

A comparative analysis of transcripts expressed in metastatic and primary breast tissues from the same individual was performed by Serial Analysis of Gene Expression ("SAGE") (U.S. Patent No. 5,695,937). Briefly, the SAGE analysis began with providing complementary deoxyribonucleic acid (cDNA) from (1) the metastatic population and (2) non-metastatic population of cells. cDNAs derived from both cell populations were linked to primer sites. Sequence tags were then created, for example, using the appropriate primers to amplify the DNA. By measuring the differences in these tags between the two cell populations, sequences which are preferentially expressed in one but not the other cell type were identified.

For the purpose of saving paper,
pages 57-157 have not be printed.

TGACTGTCAC			3718
AACCCAGGAG	Human clone 23618 mRNA sequence.	AF0071	3719
CGCACCATTG	GCN5-like 1=GCN5 homolog/putative regulator of tra	S82447	3720
TGAAAAGCTT	N8=tumor expression-enhanced gene [human, NCI H-69	S82081	3721
TCTCTACTCT			3722
TCCGTGGTTG	Homo sapiens neuronal tissue-enriched acidic prote	AF0396	3723
GTGGCGGGCG	Homo sapiens malignancy-associated protein mRNA, p	AF0414	3724
TATTTTGTTA	Homo sapiens cdc14 homolog mRNA, complete cds.	AF0003	3725
TGTCCTGGTT	Human wild-type p53 activated fragment-1 (WAF1) mR	U03106	3726
TATTTTCTTT	H.sapiens polyA site DNA sequence.	Z24749	3727
TTTCAGAGAG	Homo sapiens signal recognition particle subunit 9	U20998	3728
TGCCCTTCAA			3729
GTCTATGCCT			3730
TACTGGCCGC			3731
TATTTAAACA			3732
GCTTTTTAGA	Human non-histone chromosomal protein HMG-14 mRNA,	J02621	3733
TAGCTGTCTT			3734
ATTCTGTCAA			3735
TATCTGCCAA			3736
AAGAAGCAAG			3737
ATGGTTCTCA			3738
ATTACAGCCA			3739
ATTAACTTAT			3740
TCAGTACAGA			3741
TCAGTTCTTG			3742
ATGTCTTTTC	Human insulin-like growth factor binding protein 4	M62403	3743
TGCTGTGCAT	Homo sapiens dead box, X isoform (DBX) mRNA, alter	AF0009	3744
TCAGAAGTTT			3745
TCCTTGGACC	Human proline dehydrogenase/proline oxidase (PRODH	U82381	3746
ATTGATCAAT			3747
TTGTCCATAT			3748
TCTGCGCATC			3749
GGAGGCCGAG			3750
ATAAAACATT			3751
ATAATAAAAG	Human cytokine (GRO-gamma)	M36821	3752

CLAIMS

1. An isolated population of polynucleotides comprising or corresponding to at least one polynucleotide selected from the group consisting of SEQ ID NOS. 1 through 5911 and their respective complements.

2. A population of polynucleotides comprising or corresponding to a population of tags selected from the group 1-5, 1-17, 18-24, 1-24, 25-36, 1-36, 18-36, 37-53, 54-74, 37-74, 1-53, 1-74, 75-116, 1-116, 117-279, 1-279, 280-549, 1-549, 550-1160, 1-1160, 1161-3175, 1-3175, 3176-3183, 3184-3197, 3176-3197, 3198-3204, 3176-3204, 3205-3213, 3176-3213, 3214-3226, 3176-3226, 3227-3242, 3176-3242, 3243-3294-3176-3294, 3295-3381, 3176-3381, 3382-3554, 3176-3354, 3555-4012, 3176-4012, 4013-5911-3176-5911, 1-5911, or any combination thereof.

3. The population of claim 1, wherein the one polynucleotide comprises or corresponds to a novel tag or its complement.

4. The population of claim 1, wherein the one polynucleotide comprises or corresponds to a tag or its complement that is overexpressed in cells derived from a primary breast tumor.

5. The complement of the polynucleotide of claims 1 or 2.

6. An isolated novel polypeptide expressed by a polynucleotide of claim 5.

7. A solid phase support comprising a polynucleotide of claims 1 or 2,

8. An array of probes comprising a polynucleotide of claims 1 or 2 bound to a chip.

9. A method of aiding in the diagnoses of the metastatic condition of a metastatic breast cell comprising determining differential expression of a polynucleotide of claims 1 or 2, or the encoded polypeptide.

10. A method of modulating the genotype of a breast cell, comprising introducing into the breast cell a polynucleotide of claim 1.

11. A method of screening for a candidate therapeutic agent that modulates the expression of a polynucleotide associated the metastatic condition of a breast cell, comprising contacting a cell with an effective amount of a potential agent, and assaying for a change in expression level of a polynucleotide of claims 1 or 2, wherein a change in the expression level is indicative of a candidate therapeutic agent.

12. A polynucleotide comprising a promoter sequence derived from a polynucleotide of claim 1.

13. A host cell comprising the polynucleotide of claim 1 or 12.

14. A gene delivery vechicle comprising a polynucleotide of claim 1 or 12.

15. A polynucleotide of claim 12 and a second polynucleotide operatively linked thereto.

16. A polynucleotide of claim 15, wherein the second polynucleotide encodes an antigenic peptide.

17. A method for inducing an immune response in a subject comprising administering an effective amount of the polynucleotide of claim 1, 12 or 16, to the subject.

DECLARATION OF NON-ESTABLISHMENT OF INTERNATIONAL SEARCH REPORT

International application No.
PCT/US99/13647

The International Patent Classification (IPC) or National Classification and IPC are as listed below:

IPC(5): C07H 21/00; C12Q 1/68; A61K 48/00
US Cl.: 536/23.1, 24.3; 435/6, 320.1, 325; 530/350

4. Further Comments (Continued):

Claims 1-17 are directed to polynucleotides "comprising or corresponding to" specific decanucleotides set forth by SEQ ID NO, and the application does not comply with the requirements regarding nucleotide sequence disclosures. Claim 6 is directed to polypeptides encoded by the polynucleotides and the description does not disclose such polypeptides. Claims 12-17 are directed to promoters derived from the polynucleotides and the description does not disclose such polypeptides. The polynucleotides disclosed are only decanucleotides, which are incapable of encoding a polypeptide or serving as a promoter.